Operational Duality between Gelfand-Pinsker and Wyner-Ziv Coding

Ankit Gupta
Samsung Telecommunications America
1301 E Lookout Dr
Richardson TX 75082
agupta2@sta.samsung.com

Sergio Verdú
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544
verdu@princeton.edu

***Abstract*—We explore the duality between the Gelfand-Pinsker problem of channel coding with side information at the transmitter and the Wyner-Ziv problem of lossy compression with side information at the decompressor in the operational sense: whether a capacity-achieving encoder-decoder sequence achieves the rate distortion function of the dual problem when the channel decoder (resp. encoder) is the source compressor (resp. decompressor). We show that there exist capacity-achieving channel coding schemes that also achieve the rate-distortion function for the dual problem. However, this duality does not hold for every capacity-achieving channel coding scheme. In particular, we show that the original capacity-achieving encoder-decoder scheme of Gelfand-Pinsker operates far from the Wyner-Ziv rate-distortion function.**

I. Introduction

Duality between the problems of channel coding and lossy compression has been recognized since [1], wherein Shannon remarked: "There is a curious and provocative duality between the properties of a source with a distortion measure and those of a channel". At least three notions of duality viz. formula, functional and operational are found in the literature. These notions of duality have proved crucial towards solving various problems in information theory and designing practical coding systems. For example, functional duality has provided numerical optimization algorithms for the channel capacity and rate distortion functions [2], [3], while operational duality has been the motivation for designing lossy compressors [4] based on channel codes [5], and viceversa [6] and [7].

Formula duality refers to the relationship between the optimization problems corresponding to solving for the channel capacity and the rate-distortion function. In [8] it was shown that for the memoryless channel $P^*_{Y|X,S_1,S_2}$ with side information S_1 at the encoder and side information S_2 at the decoder jointly distributed according to $P^*_{S_1,S_2}$, and a channel cost function[1] $c(\cdot,\cdot) : \mathcal{X} \times \mathcal{S}_1 \to \mathbb{R}$, the capacity is given as

$$\bar{C}(W) = \max_{\substack{P_{U|S_1} \\ X=f(U,S_1) \\ \mathbb{E}[c(X,S_1)]\le W}} [I(U;S_2,Y) - I(U;S_1)], \quad (1)$$

where $f(\cdot,\cdot) : \mathcal{U} \times \mathcal{S}_1 \to \mathcal{X}$, and U is an auxiliary random variable, such that $|\mathcal{U}| \le |\mathcal{X}| + |\mathcal{S}|$. Analogously for lossy compression of source Y', with joint distribution P^*_{Y',S'_1,S'_2}, where side information S'_1 is available at the decompressor and S'_2 at the compressor, and the reproduction alphabet is $\mathcal{X}'$ the rate distortion function is given as

$$\bar{R}(D) = \min_{\substack{P_{U'|Y',S'_2} \\ X'=f'(U',S'_1) \\ \mathbb{E}[d(Y',X')]\le D}} [I(U';S'_2,Y') - I(U';S'_1)], \quad (2)$$

where $f'(\cdot,\cdot) : \mathcal{U}' \times \mathcal{S}'_1 \to \mathcal{X}'$, and U' is an auxiliary random variable such that $|\mathcal{U}'| \le |\mathcal{X}'| + 1$. Together (1) and (2) cover all the scenarios where (possibly noisy) side information is available to either encoder, decoder, or both. The defining equations for capacity and rate distortion function make the formula-level duality apparent.

We next consider *functional duality*. We say that the pair of problems given in (1) and (2) are functionally dual if $\mathcal{Y} = \mathcal{Y}'$, $\mathcal{X} = \mathcal{X}'$, $\mathcal{S}_1 = \mathcal{S}'_1$ and $\mathcal{S}_2 = \mathcal{S}'_2$, and there exist a distribution $\bar{P}_{Y,U,S_1,S_2}$, and a mapping $\bar{f}(\cdot,\cdot) : \mathcal{S} \times \mathcal{U} \to \mathcal{X}$, that simultaneously optimize both problems. In such a case it follows that $\bar{C}(W) = \bar{R}(D)$. In the four special cases where side information is available for channel coding (lossy compression) at the encoder (decompressor) or decoder (compressor) or both or none, conditions on the optimization problems corresponding to the capacity and rate-distortion functions were identified in [9], such that functional duality holds.

Operational duality takes a pair of functionally dual problems as its starting point. For the the problem considered in (1) a channel encoder with blocklength n, and rate R, is a mapping: $\{0,1\}^{nR} \times \mathcal{S}_1^n \to \mathcal{X}^n$, while a channel decoder is a mapping: $\mathcal{Y}^n \times \mathcal{S}_2^n \to \{0,1\}^{nR}$. For the functionally dual problem, a lossy compressor with blocklength n and rate R is a mapping: $\mathcal{Y}^n \times \mathcal{S}_2^n \to \{0,1\}^{nR}$, and a lossy decompressor is a mapping: $\{0,1\}^{nR} \times \mathcal{S}_1^n \to \mathcal{X}^n$. Thus, a channel decoder can be used as a lossy compressor and the channel encoder can be used as a lossy decompressor and viceversa. It is natural to enquire whether an optimal scheme for one problem remains optimal when used for the functionally dual problem. In [10], we explored the scenario without side information: we showed that with overwhelming probability a codebook with symbols randomly selected from the capacity-achieving distribution and the maximum likelihood channel decoder achieve the rate-

[1]We denote the alphabet of random variable Z as $\mathcal{Z}$.

distortion function for the dual problem. However, we also showed in [10] that this operational duality does not hold in general: there exist capacity-achieving encoders and decoders that operate far from the rate-distortion function when used for the dual lossy compression problem. In fact, for any capacity achieving channel codebook, there exists a decoder which preserves optimality for channel coding, while rendering it completely useless for lossy compression . Strengthening the result in [11], we also constructed a maximal code in [10], such that the probability of error for channel coding ϵ, and the probability of exceeding the desired distortion for lossy compression $\hat{\epsilon}$ satisfy

$$\epsilon < 1 - \hat{\epsilon}, \tag{3}$$

for sufficiently large blocklengths. Thus, choosing a large probability of error for channel coding leads to small probability of error for lossy compression and vice versa. In [12] a similar construction is shown for the Wyner-Ziv/Gelfand-Pinsker problems by means of a maximal nested codebook, with parameters ϵ_1 and ϵ_2. Denoting the probability of error for a pair of functionally dual Gelfand-Pinsker and Wyner-Ziv problems as ϵ and $\hat{\epsilon}$, this construction achieves

$$\epsilon \leq (1 - \epsilon_1) + \sqrt[4]{\epsilon_2}. \tag{4}$$

and

$$\hat{\epsilon} \leq (1 - \epsilon_2) + \sqrt[4]{\epsilon_1}. \tag{5}$$

This result is analogous to (3), in that choosing a large probability of channel error, we can obtain small probability of error for lossy compression, and viceversa.

Establishing operational duality allows us obtain optimal schemes for one problem by simply using optimal schemes for the other. For example, the operational duality between channel coding and lossless compression, established in [13] and [14], was used to obtain lossless compressors from capacity-achieving error correcting codes in [14]. It can also be used as a proof technique to show achievability results for one problem using coding schemes for the other [11]. Recently, operational duality has been a motivation behind using sparse graph codes for lossy compression (for e.g. [15], [16]).

In this paper we focus on the case where side information is available only at the channel encoder [17] and only at the lossy decompressor [18]. We show that the capacity achieving encoder/decoder sequence presented in the Gelfand-Pinsker paper [17] performs poorly for the dual lossy compression problem. In fact it achieves average distortion no better than a lossy compressor operating at rate 0. On the other hand, we show the existence of a sequence of encoders and decoders that operate near the fundamental limits for both problems simultaneously.

The rest of the paper is organized as follows. In Section II, we state a pair of functionally dual Gelfand-Pinkser and Wyner-Ziv problems. In Section III we present and analyze achievability schemes for these dual problems. In Section IV, we show that the capacity-achieving encoder and decoder described in [17] operates far from the fundamental limit for the functionally-dual Wyner-Ziv problem, thereby establishing that operational duality does not hold in general. A sequence of encoder and decoders that are simultaneously good for both the dual problems is presented in Section V.

II. Functionally Dual Wyner-Ziv and Gelfand-Pinsker problems

Consider the problem of coding for the channel $P_{\bar{Y}|\bar{X},\bar{S}}$ with the side information $P_{\bar{S}}$ available only at the transmitter, where, $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ denote the input, output and side information alphabets respectively. The cost constraint is denoted as $c(\cdot,\cdot) : \mathcal{X} \times \mathcal{S} \to \mathbb{R}^+$. According to [17], the capacity is given by[2]

$$C^*(W) = \max_{\substack{P_{U|\bar{S}} \\ X=f(\bar{S},U) \\ \mathbb{E}[c(X,\bar{S})]\leq W}} I(Y;U) - I(\bar{S};U) \tag{6}$$

$$= I(\bar{Y};\bar{U}) - I(\bar{S};\bar{U}), \tag{7}$$

where $f(\cdot,\cdot) : \mathcal{S} \times \mathcal{U} \to \mathcal{X}$, and U is an auxiliary random variable, taking values over an alphabet $\mathcal{U}$ such that $|\mathcal{U}| \leq |\mathcal{X}| + |\mathcal{S}|$.

Next consider the lossy compression problem with side information at the decompressor such that source and side information are jointly distributed according to $P_{\bar{Y},\bar{S}}$, and source, reconstruction and side information alphabets are $\mathcal{Y}$, $\mathcal{X}$ and $\mathcal{S}$ respectively. The rate distortion function is given by

$$R^*(D) = \min_{\substack{P_{V|\bar{Y}} \\ \hat{Y}=f'(\bar{S},V) \\ \mathbb{E}[d(\bar{Y},\hat{Y})]\leq D}} I(\bar{Y};V) - I(\bar{S};V) \tag{8}$$

$$= I(\bar{Y};\bar{V}) - I(\bar{S};\bar{V}), \tag{9}$$

where $f'(\cdot,\cdot) : \mathcal{S} \times \mathcal{V} \to \mathcal{X}$, and V is an auxiliary random variable taking values over the alphabet $\mathcal{V}$, that satisfies $|\mathcal{V}| \leq |\mathcal{X}| + 1$. Functional duality holds between the problems described in (6) and (8) if there exist a joint distribution $P_{\bar{Y},\bar{S},\bar{U}}$ and a function $\bar{f}(\cdot,\cdot) : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$ that solve both optimization problems (6) and (8). In that case, the values of $R^*(D)$ and $C^*(W)$ are identical.

III. Achievability Schemes for Gelfand-Pinsker and Wyner-Ziv problems

We now present achievability schemes for the Gelfand Pinsker channel-capacity problem and the Wyner-Ziv rate-distortion problem. These schemes are used later to show the main results of the paper in Section IV and V respectively.

A. Definitions

We will use the following notation henceforth:

- $T^n_{P_X}(\delta)$ and $T^n_{P_{X,Y}}(\delta)$ respectively denote the set of all (strongly) δ-typical n-sequences according to P_X and all (strongly) δ-jointly typical n-sequences according to $P_{X,Y}$.

[2]The dependence of the cost function upon both X and S can be motivated by the watermarking problem, wherein the distortion introduced by the watermark (X) in the message (S) must be within certain limits.

- $\mathcal{F}(n, W, \gamma)$ denotes the set of channel encoders and decoders with blocklength n, with average cost per channel use W (where the averaging is with respect to both the input message and the side information), that achieve average probability of decoding error not exceeding γ.
- $\mathcal{G}(n, D, \gamma)$ denotes the set of compressors and decompressors with blocklength n, that achieve distortion less than D with probability greater than $1-\gamma$.
- $\bar{\mathcal{G}}(n, D, \gamma)$ denotes the set of compressors and decompressors with blocklength n that achieve distortion *greater* than D with probability greater than $1-\gamma$.

We first present a result that is crucial for the analysis of the achievability schemes for the Gelfand-Pinsker and Wyner-Ziv problems. This result can be proved by a simple extension of Lemma 10.6.2 in [19]. The achievability schemes are covered in Sections III-B and III-C respectively.

Lemma 1. *Let $\mathcal{C}(n, R, U)$ denote a random codebook, with blocklength n and 2^{nR} codewords such that each symbol is chosen independently from the distribution P_U. The probability that at least one codeword is δ-jointly typical according to $P_{Y,U}$ with an i.i.d. sequence distributed according to P_Y satisfies*

$$\lim_{n\to\infty} \mathbb{P}\left[\bigcup_{c^n \in \mathcal{C}(n,R,U)} \{(Y^n, c^n) \in T^n_{P_{Y,U}}(\delta)\}\right] = \begin{cases} 0 & R < I(Y;U) - \tau(\delta) \\ 1 & R > I(Y;U) + \tau(\delta) \end{cases}, \tag{10}$$

for some $\tau(\delta)$, that vanishes as $\delta \to 0$.

B. Achievability for the Gelfand-Pinsker problem

Consider the Gelfand-Pinsker channel coding problem as stated in Section II. Fix $\epsilon > 0$ and construct the following sequence of encoders and decoders, with asymptotic rate $C^*(W) - \epsilon$.

Codebook: Construct a codebook containing $2^{\lceil n(I(\bar{Y};\bar{U})-\epsilon/2)\rceil}$ codewords of length n such that symbols are independent and identically distributed with distribution $P_{\bar{U}}$. Divide the codebook into bins each containing $2^{\lceil n(I(\bar{S};\bar{U})+\epsilon/2)\rceil}$ codewords.

Fix $\delta > 0$, and assume a lexicographic ordering on the codewords. The encoder and decoder are given as follows:

Encoder: Let $m = \{1, \cdots, 2^{\lceil n(I(\bar{Y};\bar{U})-\epsilon/2)\rceil - \lceil n(I(\bar{S};\bar{U})+\epsilon/2)\rceil}\}$ denote the message. Given a codebook C_n, side information realization s^n and message m, choose the first codeword $\mathbf{c}$ in the m^{th} bin such that $(s^n, \mathbf{c}) \in T^n_{P_{\bar{S},\bar{U}}}(\delta)$. If no such codeword exists pick the first codeword $\mathbf{c} \in C_n$. The transmitted sequence is given as

$$x^n = [\bar{f}(s_1, \mathbf{c}_1), \cdots, \bar{f}(s_n, \mathbf{c}_n)], \tag{11}$$

where $\bar{f}(\cdot,\cdot)$ is the maximizing argument in (7).

Decoder: Given a codebook C_n and a channel output realization y^n, choose the first $\mathbf{c}' \in C_n$, such that $(y^n, \mathbf{c}') \in T^n_{P_{\bar{Y},\bar{U}}}(\delta)$. Declare the index of the bin corresponding to $\mathbf{c}'$. If no such codeword exists declare the all-zero sequence.

Achievability: We denote the sequence of random codebooks constructed above as $\mathcal{C}_n$ and the sequence of encoders and decoders described above, when used with a codebook C_n as $\bar{\mathcal{E}}_{C_n}$ and $\bar{\mathcal{D}}_{C_n}$, respectively. It can be shown that (see [17])

$$\lim_{n\to\infty} \mathbb{P}[(\bar{\mathcal{E}}_{\mathcal{C}_n}, \bar{\mathcal{D}}_{\mathcal{C}_n}) \in \mathcal{F}(n, W+\gamma, \gamma)] = 1, \tag{12}$$

for any $\gamma > 0$, and sufficiently small $\delta > 0$, where $\mathcal{F}$ is defined in Section III-A.

C. Wyner-Ziv achievability scheme

Consider the lossy source coding problem with side information as defined by (8), that is dual to the problem given in (6). Fix $\epsilon > 0$ and construct the following sequence of compressors and decompressors with asymptotic rate $R^*(D) + \epsilon$.

Codebook: Construct a codebook containing $2^{\lceil n(I(\bar{Y};\bar{U})+\epsilon/2)\rceil}$ codewords of length n such that each symbol is independent and distributed according to $P_{\bar{U}}$. Divide the codebook into bins containing $2^{\lceil n(I(\bar{S};\bar{U})-\epsilon/2)\rceil}$ codewords.

Fix $\delta > 0$, and assume a lexicographic ordering on the codewords. The compressor and decompressor work as follows:

Compressor: Given a codebook C_n and a source realization y^n, choose the first codeword $\mathbf{c} \in C_n$, such that $(y^n, \mathbf{c}) \in T^n_{P_{\bar{Y},\bar{U}}}(\delta)$. If no such codeword exists choose $\mathbf{c}$ as the first codeword in C_n. Declare the bin index of $\mathbf{c}$.

Decompressor: Given a codebook C_n and side information s^n. Let m denote the bin index. Choose the first codeword $\mathbf{c}'$ in the m^{th} bin such that $(s^n, \mathbf{c}') \in T^n_{P_{\bar{S},\bar{U}}}(\delta)$, if no such $\mathbf{c}'$ exists choose the first codeword in C_n. The reconstruction is given as

$$x_1^n = [\bar{f}(s_1, \mathbf{c}'_1), \cdots, \bar{f}(s_n, \mathbf{c}'_n)], \tag{13}$$

where $\bar{f}$ is the optimizing function in (9).

Achievability: Denote the random codebook described above as $\mathcal{C}'_n$, and its k^{th} bin as $\mathcal{C}'_n(k)$. Denote the decompressor and compressor when used with a codebook C_n as $\hat{\mathcal{E}}_{C_n}$ and $\hat{\mathcal{D}}_{C_n}$ respectively. It can be shown that (see [18])

$$\lim_{n\to\infty} \mathbb{P}[(\hat{\mathcal{E}}_{\mathcal{C}'_n}, \hat{\mathcal{D}}_{\mathcal{C}'_n}) \in \mathcal{G}(n, D+\gamma, \gamma)] = 1, \tag{14}$$

for any $\gamma > 0$, and sufficiently small $\delta > 0$, where $\mathcal{G}$ is defined in Section III-A.

IV. Gelfand-Pinsker achievability scheme fails for lossy compression

In this section we consider the encoder and decoder described in Section III-B to show the achievability of channel capacity with side information at the encoder. We show that with overwhelming probability when used for the dual lossy compression problem the random encoder and decoder perform as poorly as a lossy compression system with zero rate. More formally:

Theorem 1. *Consider a pair of dual channel coding and lossy compression problems as given in Section II and let $\bar{\mathcal{E}}_{\mathcal{C}_n}$, $\bar{\mathcal{D}}_{\mathcal{C}_n}$ denote the capacity achieving random encoder and decoder described in Section III-B. Then*

$$\lim_{n\to\infty} \mathbb{P}[(\bar{\mathcal{E}}_{\mathcal{C}_n}, \bar{\mathcal{D}}_{\mathcal{C}_n}) \in \bar{\mathcal{G}}(n, D^*(0), \gamma)] = 1, \tag{15}$$

for all $\gamma > 0$, where $D^*(0)$, is the Wyner-Ziv distortion-rate function evaluated at zero rate.

Proof: Let $\mathcal{D}_n^0$ denote the lossy compressor that outputs a sequence of all zeros irrespective of the input and define

$$G \triangleq \{(y^n, C_n) : \bar{\mathcal{D}}_{C_n}(y^n) = \mathbf{0}\}, \quad (16)$$

where $\bar{\mathcal{D}}_{C_n}$ is the decoder described in Section III-B, when used with the codebook C_n, and $\mathbf{0}$ is the all zero sequence. From Lemma 1, we have

$$\mathbb{P}[(\bar{Y}^n, \mathcal{C}_n) \in G] > 1 - \varsigma, \quad (17)$$

for any $\varsigma > 0$, for all $n > n_0(\varsigma)$. From the definition of the Wyner-Ziv rate-distortion function we have

$$\mathbb{P}[d^n(\bar{Y}^n, \bar{\mathcal{E}}_{\mathcal{C}_n}(\mathcal{D}_n^0(\bar{Y}^n), \bar{S}^n)) \geq D^*(0)] > 1 - \tau, \quad (18)$$

for any $\tau > 0$ and $n > n_1(\tau)$, because the lossy compression system with $\mathcal{D}_n^0(\cdot)$ as a compressor operates at zero rate. Define

$$\begin{aligned} &f(y^n, s^n, C_n) \\ &= 1\{d^n(y^n, \bar{\mathcal{E}}_{C_n}(\bar{\mathcal{D}}_{C_n}(y^n, s^n), s^n)) \geq D^*(0)\} \\ &\quad - 1\{d^n(y^n, \bar{\mathcal{E}}_{C_n}(\mathcal{D}_n^0(y^n), s^n)) \geq D^*(0)\}. \end{aligned} \quad (19)$$

It follows that $f(y^n, s^n, C_n) = 0$, whenever $(y^n, C_n) \in G$. Therefore using (19) and (17), for $n > n_0(\varsigma)$ we have

$$\mathbb{E}[f(\bar{Y}^n, \bar{S}^n, \mathcal{C}_n)] \leq \varsigma. \quad (20)$$

Therefore

$$\mathbb{P}[d^n(\bar{Y}^n, \bar{\mathcal{E}}_{\mathcal{C}_n}(\bar{\mathcal{D}}_{\mathcal{C}_n}(Y^n), \bar{S}^n)) \geq D^*(0)] \geq 1 - \varsigma - \tau, \quad (21)$$

where (21) holds for $n > \max\{n_0(\varsigma), n_1(\tau)\}$. Since ς and τ are arbitrary we get

$$\lim_{n\to\infty} \mathbb{P}[d^n(\bar{Y}^n, \bar{\mathcal{E}}_{\mathcal{C}_n}(\bar{\mathcal{D}}_{\mathcal{C}_n}(\bar{Y}^n), \bar{S}^n)) \geq D^*(0)] = 1. \quad (22)$$

To relate (22) to (15) note that

$$\begin{aligned} &\mathbb{P}[d^n(\bar{Y}^n, \bar{\mathcal{E}}_{\mathcal{C}_n}(\bar{\mathcal{D}}_{\mathcal{C}_n}(\bar{Y}^n)), \bar{S}^n) < D^*(0)] \\ &\geq \mathbb{P}[\bar{\mathcal{D}}_{\mathcal{C}_n}, \bar{\mathcal{E}}_{\mathcal{C}_n} \in \bar{\mathcal{G}}^c(n, D^*(0), \gamma)]\gamma \end{aligned} \quad (23)$$

where (23) follows form the definition of $\mathcal{G}^c$. Taking limits we get (15), in view of (22). ■

V. Sequence of Encoders/Decoders that are optimal for both problems

In Section IV we showed that the sequence of encoders/decoders described in Section III-B to show the achievability of channel capacity with side information operate poorly as a sequence of lossy compressors and decompressors for the dual problem. In this section we describe a sequence of encoders and decoders are are optimal for both the dual problems presented in Section II. We first provide an intuitive sketch of the proof. The traditional Gelfand-Pinsker coding scheme fails for Wyner-Ziv coding, because the number of codewords in the Gelfand-Pinsker codebook is not enough for a random sequence generated from the source distribution $P_{\bar{Y}^n}$ to be jointly typical with any of the codewords. For the typical codeword search to succeed, the number of codewords in the codebook must be greater than $2^{n(I(\bar{Y};\bar{U})+\epsilon)}$, which would then rule out its applicability for channel coding. It would seem that to construct a coding scheme that works for both problems is a very challenging problem because currently in the literature there exists no proof of achievability without using strongly typical codeword search.

We solve this challenge by making use of two codebooks: one codebook is designed to be used by the Gelfand-Pinsker problem, and the other codebook is designed to be optimal for Wyner-Ziv compression. The number of bins in both codebooks are the same to ensure that the encoder and decoder output the same number of bits. The Wyner-Ziv codebook is designed to only compress part of the source sequence and ignore the rest. As the rate approaches the capacity, the penalty incurred by ignoring this segment vanishes.

However, note that the encoder is just a mapping that takes in a sequence of bits and a side information sequence in $\mathcal{S}^n$ and outputs a sequence in $\mathcal{X}^n$. It cannot discover from the bits and side information sequence which of the two problems it is facing since the input statistics are the same for both problems. Similarly the decoder just takes a sequence in $\mathcal{Y}^n$ and it has no way of determining from the input which problem it faces, and thus which of the two codebooks to use.

Fortunately, can make use of the following idea: The Gelfand-Pinsker decoder fails very rarely for channel coding. However there aren't enough codewords to ensure that one of them will be jointly typical with a sequence randomly drawn from $P_{\bar{Y}^n}$. Thus a failure of the Gelfand-Pinsker decoder signals that with high probability we are facing a Wyner-Ziv problem and we apply the compressor for the Wyner Ziv problem.

How about the encoder? Here again we make use of the fact that the Wyner-Ziv decompressor utilizes a codebook, with $2^{n(I(\bar{S}:\bar{U})-\epsilon)}$ codewords in each bin. From the foregoing argument a randomly generated sequence from $P_{\bar{S}^n}$ will not be able to locate a jointly typical codeword. Thus if we are facing a Gelfand-Pinsker problem, running the Wyner-Ziv decompressor will fail and we can run the Gelfand-Pinsker encoder instead.

Putting everything together, at the encoder, we first run the Wyner-Ziv decompressor, if it fails then we call the Gelfand-Pinsker encoder. The order is reversed at the decoder we first run the Gelfand-Pinsker decoder and then run the Wyner-Ziv compressor. Note that the order of these blocks is crucial and reversing it will cause this scheme to fail.

Let us now verify that this scheme works as desired. Suppose we are facing a Gelfand-Pinsker problem, then at the encoder we will first run the Wyner-Ziv decompressor. It will fail with high probability and we will then invoke the Gelfand-Pinsker encoder. At the decoder we will run the Gelfand-Pinsker decoder and with high probability it will recover the desired bits. Similarly for a Wyner-Ziv problem, we first call the Gelfand-Pinsker decoder, which fails with high probability. Subsequently, we run the Wyner-Ziv compressor, which succeeds with high probability. At the encoder (or decompressor) we run the Wyner-Ziv decompressor which succeeds with high probability.

We now present this scheme more formally.

Theorem 2. *Fix, $\epsilon > 0$ and $\gamma > 0$. There exists a sequence of encoders and decoders $\mathcal{E}_n$, and $\mathcal{D}_n$, with asymptotic rate $R^*(D) - \epsilon$, such that for n sufficiently large*

$$(\mathcal{E}_n, \mathcal{D}_n) \in \mathcal{F}(n, W + \gamma, \gamma), \tag{24}$$

and

$$(\mathcal{E}_n, \mathcal{D}_n) \in \mathcal{G}(n, D + \theta(\epsilon), \gamma), \tag{25}$$

for some $\theta(\epsilon)$, that vanishes as $\epsilon \to 0$.

Proof: We only present the code construction here. An intuitive sketch of optimality was presented in the discussion preceding Theorem 2. For a more rigorous treatment see [20].

Code Construction: Construct two random codebooks $\mathcal{C}_{1,n}$ and $\mathcal{C}_{2,n}$ as follows:

$\mathcal{C}_{1,n}$**:** Choose the largest integer l such that

$$\begin{aligned} &\lceil l(I(\bar{Y};\bar{U}) + \epsilon/2)\rceil - \lceil l(I(\bar{S};\bar{U}) - \epsilon/2)\rceil \\ &\leq \lceil n(I(\bar{Y};\bar{U}) - \epsilon/2)\rceil - \lceil n(I(\bar{S};\bar{U}) + \epsilon/2)\rceil. \end{aligned} \tag{26}$$

Construct a codebook containing $2^{\lceil l(I(\bar{Y};\bar{U})+\epsilon/2)\rceil}$ codewords of length l each such that each symbol is independently distributed according to $P_{\bar{U}}$. Divide the codebook into bins, such that each contains $2^{\lceil l(I(\bar{S};\bar{U})-\epsilon/2)\rceil}$ codewords.

$\mathcal{C}_{2,n}$**:** Construct a codebook containing $2^{\lceil n(I(\bar{Y};\bar{U})-\epsilon/2)\rceil}$ codewords of length n each such that each symbol is independently distributed according to $P_{\bar{U}}$. Divide the codebook into bins containing $2^{\lceil n(I(\bar{S};\bar{U})+\epsilon/2)\rceil}$ codewords.

Fix $\delta > 0$. Let x_0 be a symbol in the reproduction alphabet. Assume a lexicographic ordering on the codewords. We next describe the encoder and decoder.

Encoder: Given a side information realization s^n, denote the input $\lceil n(I(\bar{Y};\bar{U})-\epsilon/2)\rceil - \lceil n(I(\bar{S};\bar{U})+\epsilon/2)\rceil$ bits as m, and let k denote the first $\lceil l(I(\bar{Y};\bar{U})+\epsilon/2)\rceil - \lceil l(I(\bar{S};\bar{U})-\epsilon/2)\rceil$ bits. Choose the first $\mathbf{c} \in \mathcal{C}_{1,n}(k)$, such that

$$(s^l, \mathbf{c}) \in T^l_{P_{\bar{S},\bar{U}}}(\delta), \tag{27}$$

where s^l denotes the first l symbols of the side information realization. Declare

$$x^n = [f(s_1, \mathbf{c}_1), \cdots, f(s_l, \mathbf{c}_l), x_0, x_0 \cdots, x_0]. \tag{28}$$

If no such $\mathbf{c}$ exists select the first $\mathbf{c} \in \mathcal{C}_{2,n}(m)$, such that

$$(s^n, \mathbf{c}) \in T^n_{P_{\bar{S},\bar{U}}}(\delta), \tag{29}$$

and declare

$$x^n = [f(s_1, \mathbf{c}_1), \cdots, f(s_n, \mathbf{c}_n)]. \tag{30}$$

If no such $\mathbf{c}$ exists declare $x^n = [x_0, x_0, \cdots, x_0]$.

Decoder: Given the output realization y^n, select the first $\mathbf{c} \in \mathcal{C}_{2,n}$ such that

$$(y^n, \mathbf{c}) \in T^n_{P_{\bar{Y},\bar{U}}}(\delta), \tag{31}$$

and declare its bin index. If no such $\mathbf{c}$ exists, choose the first $\mathbf{c} \in \mathcal{C}_{1,n}$ such that

$$(y^l, \mathbf{c}) \in T^l_{P_{\bar{Y},\bar{U}}}(\delta), \tag{32}$$

where y^l denotes the first l symbols of the output (or source) realization. Declare the bin index of $\mathbf{c}$ and append extra zeros (if required) to make the desired total number of bits equal to $t_n = \lceil n(I(\bar{Y};\bar{U}) - \epsilon/2)\rceil - \lceil n(I(\bar{S};\bar{U}) + \epsilon/2)\rceil$. If no such codeword exists declare the all-zero sequence. ■

REFERENCES

[1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Convention*, vol. 7, pp. 142–163, Mar. 1959.

[2] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[3] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[4] M. Marcellin and T. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Transactions on Communications*, vol. 38, no. 1, pp. 82–93, Jan. 1990.

[5] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets Part I: Introduction," *IEEE Communications Magazine*, vol. 25, no. 2, pp. 5–11, Feb. 1987.

[6] R. Laroia and N. Farvardin, "A Structured Fixed-Rate Vector Quantizer Derived from a Variable-Length Scalar Quantizer: Part I-Memoryless Sources," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 851–851, May 1993.

[7] R. Laroia, N. Farvardin, and S. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1044–1056, Jul. 1994.

[8] T. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1629–1638, Jun. 2002.

[9] S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1181–1203, May 2003.

[10] A. Gupta and S. Verdú, "Operational Duality between Lossy Compression and Channel Coding," *submitted to: IEEE Transactions on Information Theory*, 2009.

[11] I. Csiszar and J. Korner, *Information Theory, Coding Theorems for Discrete Memoryless Systems*. New York Academic, 1981.

[12] H. Wang and P. Viswanath, "Fixed binning schemes: an operational duality between channel and source coding problems with side information," in *Proc. 2004 IEEE International Symposium on Information Theory (ISIT 2004)*, Seattle, Washington, USA, Jul. 2004, p. 197.

[13] T. Ancheta, "Bounds and techniques for linear source coding," Ph.D. dissertation, Department of Electrical Engineering University of Notre Dame, Notre Dame, IN, 1977.

[14] G. Caire, S. Shamai, and S. Verdú, "Lossless data compression with error correcting codes," *Advances in Network Information Theory, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, American Mathematical Society*, vol. 66, pp. 263–284, 2004.

[15] E. Martinian and M. J. Wainwright, " Low-density codes achieve the rate-distortion bound," in *Proc. 2006 Data Compression Conference (DCC 2006)*, Snowbird, Utah, USA, Mar. 2006, pp. 153–162.

[16] A. Gupta and S. Verdú, "Nonlinear Sparse-Graph Codes for Lossy Compression," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 1961–1975, May 2009.

[17] S. Gelfand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[18] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[19] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. Wiley Interscience, 2006.

[20] A. Gupta and S. Verdú, "Operational Duality between Gelfand-Pinsker and Wyner-Ziv Problems," *Preprint*.